Lines: 66

What is R? A graph drawer's perspective*

Max Sondag[1], Cagatay Turkay[2] Sibylle Mohr[3], Louise Matthews[3], Kai Xu[4], and Daniel Archambault[1]

[1] Swansea University, Swansea, UK
{m.f.m.sondag,d.w.archambault}@swansea.ac.uk
[2] Warwick University, Warwick, UK
Cagatay.Turkay@warwick.ac.uk
[3] University of Glasgow, Glasgow, UK
{Sibylle.Mohr,Louise.Matthews}@glasgow.ac.uk
[4] Middlesex University, Middlesex, UK
K.Xu@mdx.ac.uk

1 Introduction

In 2020, the worldwide pandemic of COVID-19 had a profound impact on society. One of the most important metrics that is being used to investigate the effectiveness of various interventions to control the spread of the pandemic is R, the effective reproduction number. It is talked about extensively in the news, and a wide-ranging array of different interventions are put in place by governments with the aim of getting R below 1 to curb the spread of the disease. In this abstract, we will investigate what R means from a graph drawer's perspective and aim to open up interesting and relevant research avenues.

2 Defining R

To define R, we start with a temporal network [2] G of the contacts between people, for example all contacts within a city or country. Each node u represents a person, and a temporal edge $e_t = (u, v)$ represents that person u was in contact with person v at discrete time t. We overlay the disease we are interested in on G. Some nodes will be index cases, the initial cases where the disease emanates from. These nodes are exposed to the disease from outside the network (i.e. a pangolin or international travel). A person v who has become infectious at time t, has a chance to expose their neighbors over edges with a time greater than t until v is no longer infectious. After an incubation time, these neighbors become infectious in turn, and can propagate the disease further. This propagation through the network creates an infection map: A set of rooted directed trees $\mathcal{T}$, where each node v represents a person and has a value $e(v)$ which indicates when this node was exposed to the disease. Using this infection map, we can define R.

* This work was supported by UKRI EPSRC grant EP/V033670/1 and EP/V054236/1

There are two different R's that are of interest to epidemiologists. R_0 (known as, "R naught" or "basic reproduction number"), which indicates the capacity for the disease to spread when the entire population is susceptible and no interventions are in place [1], and the R commonly used by the media: R_t ("R" or "effective reproduction number"), which takes interventions and immunity into account [3,4]. The t in R_t is the time point of interest, often a day or a week, but is ultimately dependent on the characteristics of the disease and available data. In graph-theoretical terms, we can define R_t as follows. Let Y_t denote the nodes with $e(v) = t$. Let X_t be the set of children of Y_t in $\mathcal{T}$. The R_t value for a single point in time is then: $R_t^* = \frac{|X_t|}{|Y_T|}$. Due to pragmatic issues such as reporting issues and weekly updates for policymakers, R_t is often averaged over a period of time (usually 7 days for COVID) to give the final R_t value. This calculation of R_t is shown in Figure 1.

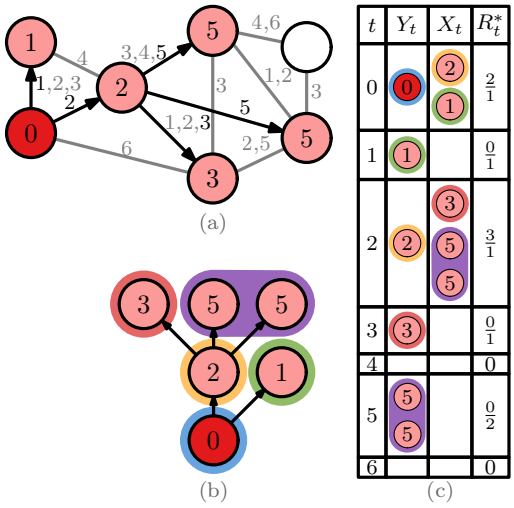

Fig. 1. Example calculation of R_t. (a) Contact graph with highlighted paths of infection. Values within nodes show exposure time $e(v)$. (b) The resulting infection map is a single infection tree. Nodes with the same exposure time have the same color. (c) R_t calculation table. R_t is the average over the 7 values of R_t^*: $(2+0+3+0+0+0+0)/7 \approx 0.71$

3 Open problems in real and simulated data

We briefly examine various (open) problems that are encountered when using this graph-based data. We first consider problems in the temporal network. In real data, both edges and nodes are missing from the graph. The time of contact could be incorrect, and is at best an approximation (often at day accuracy) even for simulated data. Continuing with problems in determining the spread of the disease over the network, the directionality of the edges is generally unknown. It is also not typically known for certain whether one nodes infected another, or if there was an outside influence. The time of exposure is often an estimation, as testing requires time, and the results might even be incorrect. For both real and simulated data, these data characteristics lead to the graph structure of the infection map being a large forest of small trees, with a few larger trees. While not all of these problems and characteristics are unique to epidemic graphs, we believe that there are nevertheless a number of interesting research questions present in this setting, and research from the Graph Drawing community could help in assisting with the current pandemic and mitigating future pandemics.

References

1. Anderson, R.M., May, R.M.: Infectious diseases of humans: dynamics and control. Oxford university press (1992)
2. Holme, P., Saramäki, J.: Temporal networks. Physics Reports **519**(3), 97 – 125 (2012)
3. Nishiura, H., Chowell, G.: The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. In: Mathematical and statistical estimation approaches in epidemiology, pp. 103–121 (2009)
4. Wallinga, J., Lipsitch, M.: How generation intervals shape the relationship between growth rates and reproductive numbers. Proceedings of the Royal Society B: Biological Sciences **274**(1609), 599–604 (2007)